INTEGRATING COMMUNITY-ENGAGED WORK INTO THE DATA SCIENCE CURRICULUM

Elizabeth Brondos Fry
St. Catherine University
ebfry704@stkate.edu

Community-engaged learning bridges course content with practical applications and allows students to apply learning in meaningful ways that contribute to social good. At St. Catherine University, community-engaged learning has been implemented in two undergraduate courses in data science and statistics at different levels. At the introductory level, students in a data visualization course used data from their community to explore long-term impacts of racially restrictive covenants that historically reserved land for the exclusive use of white people in the United States. Students in more advanced quantitative courses participated in two interdisciplinary data analysis events. In these events, students explored and analyzed large datasets in partnership with organizers in our local county to make policy recommendations for boosting economic mobility from poverty and work for racial equity in the county. This paper describes these community-engaged projects, summarizes student reception and takeaways, and provides some lessons learned and ideas for future work.

INTRODUCTION

Community-engaged learning, often referred to as service-learning, gives students experiences with applying their course-related skills to contribute to their communities, and has been identified as a high-impact practice that increases student engagement and success (Kuh, 2008). Community-engaged learning is a component of over 40 courses each semester at St. Catherine University in St. Paul, Minnesota, USA. St. Catherine University is a women's institution at the undergraduate level with a mission statement emphasizing the values of liberal arts, leadership, and social justice. Thus, community-engaged learning is a natural fit for living out the values of the institution by increasing student commitment to civic engagement, community, and social justice.

Community-engaged learning can fit in well with recommendations for student learning in statistics and data science courses. For example, the American Statistical Association's (ASA) Curriculum Guidelines for Undergraduate Programs in Statistical Science (ASA, 2014) emphasize that students should be able to work with complex data, communicate methods and results in basic terms, and make findings and visuals accessible to external audiences. Community partners who benefit from data analysis can be an external audience for students' work, in a mutually beneficial relationship that allows students to practice their skills while helping their communities. Leaders of the American Statistical Association, such as 2012 ASA President Robert Rodriguez, have advocated for using statistics to serve society, and for increasing awareness of statistics as a discipline that promotes human welfare (Rodriguez, 2013). Various instructors of statistics have incorporated service-learning experiences, not only for students of intermediate or advanced courses (e.g., Anderson & Sungur, 1999; Doehler, 2018; Gunmaratna, Johnson, and Stevens, 2007) but also for students at the introductory level (e.g., Nordmoe, 2007; Phelps, 2012).

Efforts are underway at St. Catherine University to integrate community-engaged learning into the curriculum for the statistics minor and the data science major, two relatively new programs. The university has a Center for Community Work and Learning that works to facilitate community-engaged learning in various ways, such as training faculty and finding community partners for courses. Our faculty who teach statistics and data science believe that students at all levels should be able to experience community-engaged work, rather than having to wait until capstone projects or more advanced courses. In this manner, students can experience the usefulness of statistics and data science from the very beginning, and hopefully be motivated to learn more. Currently, we have incorporated community-engaged learning as a major component of two courses: An introductory data visualization course, and a more advanced statistical computing course.

DATA VISUALIZATION AND *MAPPING PREJUDICE*

A community-engaged project was incorporated into a new introduction to data visualization course (DSCI 1000: Telling Stories with Data). The course has no pre-requisites, and although it is an

In: R Helenius, E Falck (Eds.), *Statistics Education in the Era of Data Science*
Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),
Aug-Sept 2021, Online conference. iase-web.org/Conference_Proceedings.php

entry-level course to the Data Science major, it is open to students of all majors. In the course, students learn how to craft summaries and visualizations of data guided by research on human perception, and how to interpret data visualizations to tell a story. Our hope is that the course will give students a flavor of data science, and hopefully inspire them to pursue further studies or even a major in the field. The data visualization course has only been taught once, during fall 2020. It was taught fully online (mostly synchronously) due to the coronavirus pandemic, and a total of 11 students completed the course.

One major project in this course involved community-engaged work with *Mapping Prejudice* (https://mappingprejudice.umn.edu/), a project based out of the University of Minnesota, Twin Cities that is unearthing the history of racial restrictions on buying property. In the United States during approximately the first half of the 20th century, covenants were embedded into property deeds explicitly prohibiting anybody who was not white from purchasing that property. In 1953, the Minnesota Legislature prohibited these racial restrictions in the state, but the use of covenants remained in use throughout much of the United States until 1968, when a federal law called the Fair Housing Act made them explicitly illegal. The *Mapping Prejudice* project collected thousands of housing deeds throughout Hennepin County (the county where Minneapolis, Minnesota is located) and created a map of the locations of these covenants which can be seen on their homepage. Since then, the project has expanded to create maps for other regions, including collaboration with St. Catherine University's *Welcoming the Dear Neighbor?* (https://welcomingthedearneighbor.org/) project to create a map of racial covenants for Ramsey County (the county where St. Paul, Minnesota, where St. Catherine is located) and to investigate the history and effects of racialized housing discrimination in the county.

During the fall of 2020, the map of racial covenant locations was fully constructed for Hennepin County, but *Mapping Prejudice* and *Welcoming the Dear Neighbor?* were still working with many community volunteers to read thousands of housing deeds to create a database to plot and map the covenants in Ramsey County. As part of their community-engaged learning for the Data Visualization course, students transcribed housing deeds (flagging racially restrictive language and detailing the location) along with community volunteers, while they were learning about the complexities of data collection in the course. The students spent about two hours being trained and participating in the housing deeds transcription.

The majority of the community-engaged learning for the course involved a group project that took place over most of the semester, in which students found, cleaned, and visually explored Hennepin County data to explore potential long-lasting effects of housing discrimination in the county. The final product of this project was to create a blog post with data visualizations answering the question: "Why do racial covenants still matter today?" to be shared on the *Welcoming the Dear Neighbor?* website for a broad audience interested in this project. Students also gave a (virtual) class presentation with a few external guests invited from the Center for Community Work and Learning.

This was a large project, so it was broken down into stages:

- *Stage 1: Topic Exploration.* This first stage was completed individually. Students found and summarized articles online related to impacts of housing inequality. They chose specific topics of interest (e.g., education, health, policing, air quality). The instructor used their responses to place students into groups based on topics of interest.
- *Stage 2: Data Exploration.* In their groups, students found at least two datasets (from Minneapolis, or from Hennepin County at large) related to their topic of interest and identified variables that they could use to build visuals.
- *Stage 3: Creating and Critiquing Visuals.* Students created visuals using their group's data and posted these to an online discussion board. Then, students reviewed each other's visuals and replied to each other's posts interpreting the story that they saw from each visual. In their replies, they also included constructive critiques noting the strengths of the visual and suggestions for improvement.
- *Stage 4: Final blog post and presentation.* Student groups tied together their visuals from Stage 3 into a written story about exploring the lasting impacts of racial covenants. They presented this story orally on the last day of class, and also created a written blog post to share with a broader audience online.

A total of four class periods were set aside as project workdays for students to meet virtually and work on their project stages.

In addition to content-related work, students were given the opportunity to reflect on their community-engaged experiences. Using tools provided by the university's Center for Community Work and Learning, students posted reflections about what they noticed during their experiences transcribing covenants, as well as reflections at the end of the semester about how they could continue to work for social change, using what they had learned.

STATISTICS AND ECONOMICS CLASSES: WORK WITH RAMSEY COUNTY

Community-engaged learning, partnering with the Ramsey County Community & Economic Development team, was implemented in an upper-level statistics course in collaboration with an economics course. The statistics course (STAT 3090: Statistical Computing) covers statistical learning topics and is primarily taken by statistics minors, although it is open to anyone who has taken introductory statistics and has had exposure to multiple regression and using R statistical software. I taught this class virtually during spring 2021 with 7 students enrolled. For our community-engaged work, our class teamed up with a professor and students in a Quantitative Impact Evaluation course offered by our university's Economics and Political Science department, with 17 students enrolled. In that class, students who had already taken introductory statistics learned econometric techniques and applied them to data in the social sciences.

Our community partners from Ramsey County were applying for a grant to fund actions that would increase upward mobility from poverty. Representatives from the county first met with the two course instructors and staff members from St. Catherine's Center for Community Work and Learning to talk about how students could help to carry out data analysis and report results to provide ideas for policies to boost upward mobility. A plan was made to have two data events, similar to "data marathons", where students would work in teams to explore and analyze large datasets related to poverty and upward mobility in Ramsey County. We drew inspiration from the ASA event "Data Fest" (Gould, 2014) and even borrowed the same name, although our events were smaller and shorter. Each data event began on a Tuesday afternoon around 3pm (when the Statistical Computing class typically met), then went into the evening until about 9-10pm. Students met again the following morning from 8:00-9:30am (when the Quantitative Impact Evaluation class typically met) to present their teams' results to their peers and instructors. The first data event was held during Week 4 of the 15-week semester (focused on basic exploratory data analysis), while the second event was held during Week 10 (with more of a statistical modeling focus).

The two data events were planned and organized by the two faculty members teaching the statistics and economics classes, a staff member from the Economics department lab, and two student assistants (one teaching assistant and another student taking independent study credits). Prior to the data events, the students and instructors in both classes were invited to meet with the Ramsey County partners to learn more about their work on helping to boost upward mobility. The student assistants found and prepared relevant data sets and constructed a set of tasks to scaffold students through the processes of data exploration and modeling. Data sets on intergenerational mobility were used from Opportunity Insights (https://opportunityinsights.org/) and included information at the Census tract level from a cohort of individuals born in the late 1970s and early 1980s who were then tracked later as adults in the 2010s. For example, one of the variables describing "upward mobility of the 25th percentile" was the average income, in dollars, of adults in a census tract who as children grew up in families at the 25th percentile of income distribution. Similar variables existed for people who grew up in other percentiles (e.g., 50th, 75th) and subset by gender and race. The data also contained hundreds of variables at the Census tract level, such as the median rent, the proportion of households with single parents, the share of people who are non-white, unemployment rate, and poverty rate. During the data events, students were challenged to answer overarching questions such as:

- Is Ramsey County a place that promotes upward mobility?
- What does Ramsey County need to do to develop more upward mobility?
- In what areas should they invest? How can they address inequality?

Participation in the two data events was a requirement for all students in the two statistics and economics classes, but the events were open to all students in the university community, regardless of their prior background with data analysis. In addition, announcements were e-mailed to invite recent alumni and students from a few surrounding universities. During the first event, there were a handful of outside participants, but during the second event, the participants were nearly all students in the statistics and economics classes. Prior to each data event, students took a survey indicating their comfort level with R (software used in the statistics course) and Stata (software used in the economics course). The two instructors placed students into either R or Stata teams of about 4 students each and ensured that there was a mixture of students with a variety of different comfort levels with the software tool used by that team. We also assigned an informal student mentor for each team, who was either an advanced student or alumna joining us, and their role was to provide leadership and help with software questions that arose. In addition, the professors and a few "floating" student/staff mentors checked in with the groups throughout the events. Students worked in teams during the Tuesday evening portion of the event (with multiple short breaks built in to eat and socialize) to put together a five-minute presentation to share with the class the following morning. In addition, the students in the classes were required to write short reports of their results and their own reflections after each of the two events.

After the two events, near the end of the semester, students put together a longer presentation to share with the Ramsey County partners to share their work and give ideas for future work that would be completed on this topic by summer research students. Students were also invited to complete an end-of-semester survey about their community-engaged experiences by the Center for Community Work & Learning at St. Catherine University.

STUDENT RECEPTION

I have not tested changes in cognitive or non-cognitive outcomes as a result of community-engaged learning, but can offer informal observations on how students responded to the activities. I have observed that students who normally struggle on traditional assessments (e.g., exams) have succeeded in making meaningful contributions to their group and writing successful reports. Students have shared that the community-engaged learning opportunities have provided them with an opportunity to apply their skills in a useful way, and that the experiences helped to increase their confidence. Students who are normally quiet during class tend to speak up more during small group collaborative tasks and are not afraid to ask questions of their peers or student mentors.

Students had the opportunity to reflect on the context of their community-engaged work. Their reactions were mostly positive, and support the notion that we should continue to offer these opportunities in the future. The fall Data Visualization students reflected more informally in class discussions about their experiences, whereas in the spring I incorporated written reflections for the Statistical Computing students as a part of their grade. (In the future, I hope to incorporate more written reflection for the Data Visualization class too.) In their written reflections, the students who participated in the data events commented on their successes and struggles, their primary takeaways, and how the community could benefit from their work.

In their reflections, many students expressed that although they were working under pressure during the data events, the experiences helped them to increase their knowledge, skills, and value of collaboration. A few students reported that they had been wary of group work and group projects before college, but as a result of these projects, they appreciated the opportunity to collaborate and took pride in what their groups accomplished. Although some students found it challenging to work with group members of a wide variety of skill levels, they recognized that everyone had skills to contribute and even learned group collaboration skills in the process. As one student put it, "*Before…, I may have been more hesitant about coding collaboratively…even when I had worked on code with another person, it was more a matter of working on our own and then connecting back. This was actually working through the coding together…The fact that everyone came with their own ways of doing things meant that each brought a unique perspective to the work*."

Students in all three classes (Data Visualization, Statistical Computing, and Economics Quantitative Impact Evaluation) expressed enthusiasm about being active participants in improving their community, and seeing how their skills in data analysis could be useful in a meaningful context. For example, some students wrote:

- *"There is a common misconception that statistics and data science are impersonal and purely focused on numbers. I disagree--statistics and data analysis are tools for social change, and my service-learning work was proof of that."*
- *"Our collaboration was a bit like 'co-liberation' ... working together we can make real impacts for our community."*

Although students overall felt that their work could have an impact, they were also aware of the limitations of what they could do during their limited time. As one student commented, "*Although I don't think this particular data set will provide Ramsey County with answers to their questions, I think we were able to use the data to recognize issues…we or others can then find other data [to] explore the issues more in depth*." Students recognized the limitations of the conclusions they could make with observational data, but still valued the opportunity to provide a starting point for future work that could make a real difference in the community.

FUTURE WORK

Overall, our experiences integrating community-engaged work into courses involving data analysis have been positive, and we hope to continue similar experiences in the future even as we return to classes back in person. We also hope to infuse more community-engaged experiences throughout the curriculum, starting with a new writing-intensive data science course that will be offered next spring called "Analyzing Social Issues with Data."

Throughout our experiences this past year, we have learned some lessons to take away as we continue this work in the future. For example, collecting student reflections has been valuable not only as a metacognitive exercise for the students, but also for instructors to obtain feedback about what students take away from the experiences and what could be changed. For instance, based on student feedback after our first data event, we learned that exploring a large codebook can be intimidating, and more class time could be spent preparing students and giving them time to explore the data on their own before the actual events. We have learned that providing structure and scaffolding is important so that students do not feel lost, but at the same time, it is important to be open to flexibility. For example, the second data event happened during a week when policing and racial injustice were prevalent issues in current events in our metropolitan area. Our planning team had written a set of instructions for students to explore different models for predicting upward economic mobility, but some student groups chose to shift their focus towards looking at variables related to criminal justice instead. Finally, although flexibility is important, it is also valuable to have some early deadlines for drafts so that students can get feedback to create a better final product. For the Data Visualization blog post, students critiqued their peers' visuals, but did not have time to review drafts of each other's final blog posts. The instructor and other collaborators of the "Welcoming the Dear Neighbor?" project had to do considerable copy editing before blog posts could be shared. In the future, earlier peer reviews of the blog posts will be incorporated.

Although there is anecdotal evidence that these community-engaged experiences have been beneficial for student learning and community impact, there is ample opportunity to do research to examine student learning outcomes (on both the statistics/data science content, and general skills such as collaborating with a group). Faculty collaborators in the Economics and Political Science department are planning to test their hypothesis that courses including experiences like the community-engaged data events increase feelings of relevance, belonging, and growth mindset in women and under-represented students (Bayer, Bhanot, Bronchett & O'Connell 2020). Research can also examine changes in non-cognitive outcomes such as attitudes towards statistics and data science, perceived confidence in students' own abilities, and attitudes towards service-learning and civic engagement.

SUMMARY AND CONCLUSION

Community-engaged learning has been integrated into introductory Data Visualization and more advanced courses in statistics and economics at St. Catherine University, having students explore data related to social justice issues in our local area. We have experienced that overall, the students, faculty, and community partners have found these experiences to be a meaningful way for students to apply their course knowledge to benefit the wider community, and we hope to integrate community-

engaged learning into more courses in the future. Research is needed to explore cognitive and non-cognitive impacts of these experiences on student learning.

REFERENCES

American Statistical Association Undergraduate Guidelines Workgroup (2014). *Curriculum guidelines for undergraduate programs instatistical science.* American Statistical Association, Alexandria, VA, *http://www.amstat.org/education/curriculumguidelines.cfm*

Anderson, J. E., & Sungur, E. A. (1999). Community service statistics projects. *The American Statistician, 53*(2), 132–136.

Bayer, A., Bhanot, S., Bronchetti E., & O'Connell, S. (2020). Diagnosing the Learning Environment for Diverse Students in Introductory Economics: An Analysis of Relevance, Belonging, and Growth Mindsets. *AEA Papers and Proceedings, 110:* 294-298.

Doehler, K. (2018). Successful service-learning for statistics students studying survey sampling. *Statistics Education Research Journal, 17*(2), 82-103, *https://iase-web.org/documents/SERJ/SERJ17(2)_Doehler.pdf*

Gould, R. (2014). Datafest: Celebrating data in the data deluge. In K. Makar, B. de Sousa, and R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics* (ICOTS9, July, 2014), Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute, *https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_4F2_GOULD.pdf*

Gunaratna, N. S., Johnson, C. A., & Stevens, J. R. (2007). Service-learning for graduate students through a student-run consulting program. *Journal of Statistics Education, 15*(2), *http://ww2.amstat.org/publications/jse/v15n2/gunaratna.pdf*

Kuh, G.D. (2008). *High Impact Educational Practices: What They Are, Who Has Access to Them, and Why They Matter.* Washington, D.C.: Association of American Colleges and Universities. *https://www.aacu.org/leap/hips.*

Nordmoe, E. D. (2007). Service-learning in introductory statistics at Kalamazoo College. *Journal of Statistics Education, 15*(2), *http://ww2.amstat.org/publications/jse/v15n2/nordmoe.html*

Phelps, A. L. (2012). Stepping from service-learning to SERVICE-LEARNING pedagogy. *Journal of Statistics Education, 20*(3), *http://ww2.amstat.org/publications/jse/v20n3/phelps.pdf*

Rodriguez, R. N. (2013). Building the big tent for statistics. *Journal of the American Statistical Association, 108*(501), 1–6.